



צבי צורי

מוטציות בערימה של גנים

מתמטיקה וטכנולוגיית הריצוף חוברות בשיטה חדשה למציאת מוטציות באוכלוסיות גדולות

את השיטה פיתחו במקביל ארבעה חוקרים ישראלים: ד"ר נעם שנטל מהאוניברסיטה הפתוחה, ד"ר אמנון עמיר ממכון ויצמן למדע וד"ר אור צוק ממכון Broad של MIT והרווארד, וכן ד"ר יניב ארליך מ-MIT.

לך חפש מוטציה בערימת גנים

שיטה שיודעת לזהות בקרב אוכלוסייה גדולה מי הוא נשא של מחלה גנטית מסוימת ויודעת לעשות את זה בזול היא שיטה חשובה. אבל השיטה שפיתחו ד"ר שנטל וחבריו יודעת לעשות עוד משהו, חשוב לא פחות: היא יודעת לזהות מוטציות חדשות באוכלוסייה הנבדקת. הדבר נשמע כמו חלומה של כל חוקר בתחום הביולוגיה: לגלות מתוך אלפי דיגמות צמחים באיזה צמח מופיעה מוטציה רצויה (למשל

כל הורה מכיר את החרדה המלווה את חודשי ההיריון, שמתגברת ככל שמתקרבים אל רגע הלידה. זו החרדה מפני מה ש"מתבשל" ברחם ועשוי להתגלות רק עם צאת העובר אל אוויר העולם: פגמים מולדים שיפגעו קשה בשמחת הלידה. הרפואה המודרנית מנסה למנוע הפתעות לא נעימות מן הסוג הזה בעזרת סדרה של בדיקות של נשאות גנטית להורים, אולם הבדיקות הגנטיות האלה יקרות והדבר מגביל מאוד את מספר הבדיקות שנעשות בחינם שאותן מציעה המדינה במסגרת סל הבריאות. התוצאה היא שמי שאינו נכלל ברשימת הזכאים נאלץ להוציא מכיסו כסף רב. שיטה חדשה מציעה דרך יעילה וחכמונית לבצע את הבדיקות החיוניות הללו, לקצץ את עלותן בעשרות אחוזים וכך לאפשר למדינה להציע אותן לאנשים רבים יותר.

זו הגורמת לצמיחת עלים גדולים במיוחד), מי הגן האחראי ואיפה בדיוק ממוקמת המוטציה בתוך הגן הזה. מי שמצוי אפילו מעט בענייני גנטיקה יודע כמה עבודה יש להשקיע כדי לפצח את מבנה הדנ"א של צמח וכמה מאמץ נדרש כדי לגלות מה מיקומו של גן מסוים. בוטנאי יכול היום לקחת דגימה של צמח ואם הוא רוצה לדעת איזה נוקלאוטיד נמצא בתוך גן מסוים. בעזרת מכשירי המעבדה, החוקר "מרצף" את הדגימה ומגלה את הרצף. אבל היום צריך החוקר לחזור על כך שוב ושוב, כי בכל בדיקה הוא מקבל רק תשובה אחת. אין בכך עזרה רבה, מכיוון שהוא רוצה לדעת איפה בדיוק קורה הדבר הנדיר שהוא מחפש ב-10,000 הדגימות שלפניו. ואין מדובר רק בטרחה, אלא כמובן גם בכסף מפני שכל ריצוף עולה לא מעט. כך שלא רק הורים אלא גם ביולוגים ישמחו מאוד לאפשרות לעשות את הבדיקה הגנטית הנחוצה במהירות ובחיסכון כספי ניכר.

הקסם המתמטי מספק את התשובה

ד"ר שנטל הוא חבר סגל אורח באוניברסיטה הפתוחה ואיש המחלקה למדעי המחשב המסייע בהקמת מחלקה חדשה למדעי הקוגניציה באוניברסיטה זו. מסביר ד"ר שנטל: "הגישה הפשוטה ביותר לזיהוי נשאים היא כמובן לבדוק כל פרט באוכלוסייה. התוצאה במקרה זה תהיה חד-משמעית אבל העלות בלתי אפשרית. השיטה שלנו פועלת בדרך אחרת. מתוך אוכלוסייה מסוימת, נאמר של עשרה אנשים, אני לוקח דגימות דנ"א ממחציתם, מערבב את הדגימות של כולם בתוך מבחנה אחת, עורך את הבדיקה הנדרשת ומקבל מספר מסוים, שהוא תוצאת הבדיקה על פי ה'שיטה'." "לאחר מכן אני לוקח מדגם אקראי נוסף מתוך אותה אוכלוסייה, מערבב את דגימות הדנ"א של הנבחרים האלה (מהם שהיו כבר בקבוצת הנבדקים הראשונה ומהם שלא היו) בתוך מבחנה אחת, בודק את תוכנה ומקבל על פי השיטה מספר שני. אני עושה זאת שוב פעם שלישית, ומקפיד שכל אחד מעשרת הפריטים של האוכלוסייה הנבדקת יופיע לפחות פעם אחת באחד משלושת המדגמים האקראיים שערכתי, ומערבב שוב את הדגימות. בסוף הבדיקה מתקבל מספר שליש. בעזרת שלושת המספרים – שהם תוצאות הבדיקה הגנטית של המבחנות המעורבות – השיטה מפעילה אלגוריתם מתמטי שנותן לנו זיהוי גנטי מדויק ואישי של עשרת הפרטים באוכלוסייה הנבדקת ותשובה חד-משמעית מי נשא ומי אינו נשא. כלומר, במקום עשר בדיקות פרטיות מקבלים זיהוי חד-

משמעי בתוך מספר בדיקות נמוך בהרבה מהמקובל." לדברי ד"ר שנטל, השיטה שפיתחו הוא ושותפיו היא פרי הניצול המשולב של שני חידושים מדעיים השייכים לשני תחומים שונים. מצד אחד טכניקה חדשה מאוד של ריצוף הדנ"א, שפותחה בשנותיים האחרונות וידועה בשם הדור הבא של טכנולוגיית הריצוף (Next Generation Sequencing Technology), ומצד אחר תיאוריה מתמטית חדשה, מחמש השנים האחרונות, בשם חישה מתומצמת (Compressed Sensing).

מערבוב אקראי לזיהוי חד-ערכי

הריצוף של הדנ"א בפרויקט הגנום האנושי נעשה בשיטת סנגר, על שמו של הכימאי הבריטי פרדריק סנגר (Sanger) שהמציא את השיטה וקיבל על כך פרס נובל. מאז פותחו שיטות וטכנולוגיות חדשות המאפשרות לרצף את כל הגנום או חלקים ממנו ביעילות רבה יותר. אבל המתמטיקה היא שאחראית על ה"פלא" שמחוללת השיטה המצליחה, על סמך הדגימות המעורבות, לשחזר את הדגימה האישית של כל פרט ופרט וכך לגלות מי נשא ומי לא. "המתמטיקה החדשה היא שמשכילה לדלות את הנתונים המדויקים מתוך אותם מספרים, שהם תוצאות הערבוב של הדגימות האישיות, ומן האופן שבו בחרנו את הדגימות שעורבבו, כלומר הרכב הדגימות הפרטיות שהשתתפו בדגימות המשותפות", אומר שנטל. עוד עולה שהבחירה האקראית של הפרטים שדגימותיהם משתתפות בדגימות המעורבות אינה מכשול, להפך. המתמטיקה אומרת שכל שהבחירה אקראית יותר, כך גדל הסיכוי לקבל בשחזור תשובה מדויקת יותר.

את המתמטיקה שמאחורי השיטה החדשה פיתחו מתמטיקאים מהשורה הראשונה בעולם, החוקרים דיוויד דונוהו, עמנואל קנדס וטרנס טאו (Donoho, Candes, Tao), האבות המייסדים של התחום והאחראים לשתי פריצות דרך במתמטיקה: האחת היא אופן הערבוב של הדגימות. השלושה הראו שלערבוב אקראי יש תכונות המאפשרות שחזור מדויק דווקא של הדגימה האישית מתוך הדגימה המעורבת. פריצת הדרך השנייה קשורה בפתרון בעיית האופטימיזציה של השחזור. השלושה הראו כי אפשר לתרגם את הבעיה המקורית, שאינה ניתנת לפתרון, לבעיה אחרת שאותה אפשר לפתור בקלות. אם מערבבים את הדגימות בדרך שהציעו, מובטח שהשחזור ייתן את הפתרון הנכון בסבירות גבוהה מאוד.



השיטה מתאימה לבדיקות סקר למציאת נשאים של מוטציות ידועות, וכן למציאת מוטציות נדירות חדשות בנוסף לאלה הידועות, למשל בסיס גנטי לאוטיזם

חדשנות הרעיון

"המאמר שלנו התפרסם בכתב-העת *Nucleic Acids Research*, שמפרסם בדרך כלל רק עבודות שהוכחו בניסוי ממשי ולא הודגמו אך ורק בסימולציות מחשב", מספר שנטל. "במקרה שלנו חרג כתב-העת מהכלל הזה בשל חדשנות הרעיון. באותו שלב כל מה שהיה לנו ביד היה לא יותר מהצעה, משהו שטרם נבדק בניסוי. מאז עשינו כברת דרך וכיום השיטה כבר נמצאת בבדיקה מעשית. התוצאות הראשונות מעודדות מאוד ומאשרות את תקפותה של השיטה."

קיים קשר בין שכיחותה של מוטציה באוכלוסייה למספר הדגימות שיש לבדוק כדי לקבל, לפי שיטתם של ד"ר שנטל ושותפיו, תשובות על כל פרט ופרט באוכלוסייה. אם, למשל, שכיחות המוטציה באוכלוסייה היא 0.1%, אזי 200 בדיקות של דגימות מעורבבות, שבהן מיוצגים כל הפרטים באוכלוסייה, יספיקו כדי לגלות את כל הנשאים באוכלוסייה של 10,000 איש. אבל אם השכיחות היא של 1%, אזי 200 בדיקות יספיקו בעבור 2,000 איש בלבד. ככל שהמוטציה נדירה יותר השיטה יעילה יותר. אם שכיחות המוטציה מגיעה לכדי 5% השיטה מפסיקה להיות כדאית ומשתווה לדרכים המקובלות והשמרניות לגילוי. אלא שהאתגר המדעי נמצא דווקא במוטציות הנדירות. במקרים האחרים, הנפוצים, כבר

נמצאו דרכים יעילות לגלות את הנשאים. הבעיה שנותרה היא במקרה של המחלות הלא שכיחות ושל המוטציות הנדירות, שם מתגלה כאמור כוחה ויעילותה של השיטה החדשה. למעשה, השיטה מתאימה לשני סוגים של בעיות שעשויות להטריד הורים: הראשונה היא בדיקות סקר למציאת נשאים של מוטציות ידועות – בדיקות הריון או כל בדיקת סקר גנטית אחרת. השנייה היא מציאת מוטציות נדירות חדשות בנוסף לאלה הידועות, כפי שנעשה במחקרים (association studies) המנסים למצוא בסיס גנטי של תופעות, למשל בסיס גנטי לאוטיזם, לסכיזופרניה ועוד.

השיטה החדשה היא שיטה של "בדיקה קבוצתית" (group testing), שאותה נקט צבא ארצות-הברית כבר במלחמת העולם השנייה כשביקש לזהות בקרב חייליו חולי עגבת. במקום לבדוק את דמו של כל מתגייס, נלקחו דגימות דם מחמישה חיילים ובדקו את התערובת. אם זו התגלתה נקייה, עברו אל החמישייה הבאה ורק כאשר נמצא סימן לעגבת בתערובת נבדק כל פרט ופרט לחוד. כך חסכו בדיקות רבות. גם בשיטה החדשה מערבבים דגימות פרטיות, אבל בגלל המתמטיקה המתקדמת שבה משתמשים שוב אין צורך לחזור לדגימה האישית של כל פרט, משום שהשיטה מצליחה לשחזר בעצמה את הדגימה האישית מתוך תוצאות הדגימות המערבבות, וזאת בזכות תיאוריית החישה המתומצתת.

ניסוי בקהילה החרדית בברוקלין

השיטה החדשה נכנסה השנה לשלב הניסוי בארצות-הברית. ד"ר יניב ארליך מ-MIT יצר קשר עם אגודה חרדית בברוקלין בשם "דור ישרים", הנוהגת לערוך בדיקות גנטיות בקרב נערים ונערות חרדיים בטרם הגיעו לפרקם ולגלות עוד לפני השידוך אם קשר הנישואים עלול להיות בעייתי מבחינה גנטית (חרדים אינם עורכים בדיקות גנטיות בהיריון, מאחר שהפסקת היריון יזומה אינה אפשרות מבחינתם). נתוני הבדיקות נשמרים במשרדי הקהילה ובני המשפחות יכולים לבדוק את התאמתם הגנטית של שני הצעירים המיועדים להינשא.

כעת מבקשת הקהילה לייעל את השיטה ובמקום לערוך 3,000 בדיקות אישיות לגילוי נשאות לטיי-זקס, או לכל מחלה גנטית אחרת, היא רוצה להגיע לאותה מידה של זיהוי תוך כדי ביצוע מספר קטן יותר של בדיקות. השיטה החדשה, שיטת שנטל-עמיר-צוק-ארליך (שלה קראו החוקרים גם בשם ComSeq, קיצור של compressed sequencing), מבטיחה לעשות זאת. כך נכנסה השיטה לשלב הניסוי המעשי.

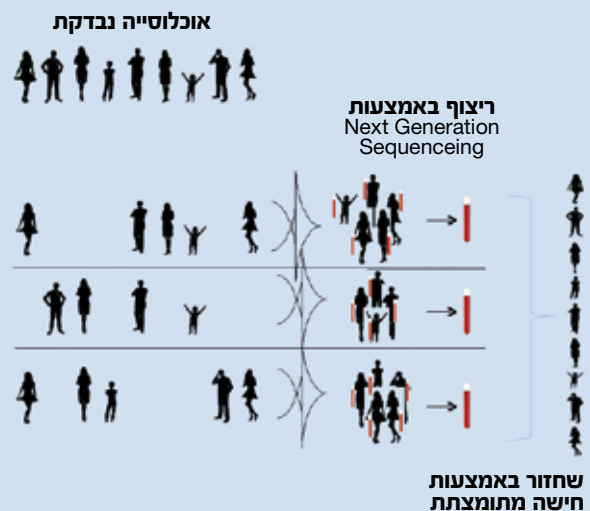
מיוחדות, למשל עמידות למליחות קרקע, או זנים בעלי מספר רב יותר של גרעינים וכו'. על כל תכונה כזאת אחראים גנים, והדבר היחיד שיכול להבטיח שינוי בתכונה הוא מוטציה. אפשר לחולל מוטציה יזומה בגן מסוים, אך במקרה כזה לעולם אין לדעת מראש מה תהיה השפעתה של המוטציה שחוללת. כל מה שאפשר לעשות הוא לחולל מספר רב של מוטציות ולקוות שלפחות אחת מהן תיתן את השינוי הרצוי. אומר שנטל: "בשיטה המקובלת כיום חושפים את הצמח לחומר המחולל מוטציות אקראיות ומחפשים בין אלפי צמחים את הצמח או את הצמחים שבהם חוללה המוטציה את השדרוג המבוקש. כלומר, היכן המוטציה הביאה לכך שמספר הזרעים יהיה גדול יותר, והיכן היא הגדילה את שטח העלים. בשלב הבא מרצפים את כל הגנים המעניינים, בכל אלפי הצמחים, כדי להבין את הבסיס הגנטי של התופעה, פעולה הכרוכה במשאבים רבים. השיטה שלנו אמורה למצוא את כל המוטציות בכל הצמחים שהשתתפו בניסוי ובכך תאפשר להבין את התופעה. בניסוי זה יהיו בידינו דגימות של 6,400 צמחים שונים, ונחפש את הבסיס הגנטי של בקרת משטר המים בצמח, אך נוכל גם לשאול כל שאלה מדעית אחרת לגבי אוכלוסיית הצמחים. זה הניסוי השני שבו עומדת להיבחן השיטה החדשה, במקביל לניסוי שנערך בדגימות הדם של עמותת דור ישרים." ❖

לקריאה נוספת

Identification of rare alleles and their carriers using compressed se(que)nsing, Shental N., Amir A. and Zuk O., Nucleic Acids Research, 38(19):e179, p. 1-22, 2010.

Compressed Genotyping, Erlich, Y.; Gordon, A.; Brand, M.; Hannon, G.J.; Mitra, P.P.; IEEE Information Theory, 56(2), p. 706-723, 2010.

Compressed Sensing Approaches for High Throughput Carrier Screen, Erlich Y., Shental N., Amir A. and Zuk O., 47th Annual Allerton Conference on Communication, Control, and Computing, p. 539-544, 2009.



שחזור באמצעות שיחה מתומצתת. איור: ד"ר נעם שנטל

ד"ר שנטל: "בידינו 3,000 דגימות דנ"א שכבר נבדקו וידוע מי מבין ה-3,000 נשא ומי לא. אנחנו ניקח את 3,000 הדגימות האלה ונערבב מהן ערבובים אקראיים. על סמך 100 או 200 דגימות מעורבבות, שבהן נמצא דם מכל 3,000 הפרטים, ננסה לזהות מי נשא ומי לא; כך נוכל להשוות בין התוצאות שלנו לנתוני האמת שבידי עמותת דור ישרים. עוד לא ברור כמה בדיקות נצטרך לעשות באותן 3,000 דגימות. אני מניח שכ-200, אך הדבר תלוי בשכיחות המוטציה. אם היא נדירה יותר נצטרך פחות בדיקות, ואם היא שכיחה יותר נצטרך יותר. יש דרך להעריך די מהר כמה בדיקות נצטרך ואם צריך להוסיף בדיקות בגלל השכיחות הגבוהה של המוטציה. אני מעריך שנוכל להקטין בכ-90 אחוז את ההוצאות, וזה לא מעט. אם נגיע לחיסכון כזה, הדבר יגדיל מאוד את הסיכוי למצוא מימון למחקר."

שימושים נוספים לשיטה

כיוון אחר שבו נבדקת השיטה החדשה הוא בגנטיקה של צמחים. ד"ר שנטל, ד"ר איל פרידמן מהפקולטה לחקלאות באוניברסיטה העברית, ד"ר ארליך וחוקר נוסף מהמכון למחקר חקלאי בארצות-הברית גיבשו הצעת מחקר במטרה להראות שהשיטה יעילה גם לפיתוח זנים מסוימים של דורה (*Sorghum bicolor*), שהוא אחד מחמשת הדגנים החשובים בעולם ומשמש מקור מזון לבני אדם ולבהמות, בעיקר באזורים צחיחים דוגמת אפריקה. יש חוקרים שמעוניינים לפתח זני דורה בעלי תכונות